VI.1 Creation of the Data Descriptor

The creation of the Data Descriptor is the first step in the process of creating a time series directory for a survey included in the Time Series Analytic Repository (TSAR) system. The Data Descriptor contains the data definitions for all time series associated with a survey. Thus, the Data Descriptor contains data definitions for many series. There is a separate record in the Data Descriptor file for each series associated with a survey. The descriptor file contains logical definitions for each time series. The time series data files themselves include the physical definition in the file header.

The function "loaddesc.mac" is the SAS MACRO that TSAR uses to create the Data Descriptor based upon inputs provided to the function. The algorithm for this MACRO is quite simple. This function reads a fixed length ASCII data file that contains data definition information for all time series associated with a survey and then writes the Data Descriptor. The system user is responsible for developing the ASCII input file.

Input to this function is a fixed length ASCII file that contains one (1) or more records, where each record defines a time series in the data repository for a survey. The input file is a fixed length file format of 761 bytes (ASCII characters) per record. Each record is delimited with an end-of-record character.

This input ASCII data set must be transferred to the UNIX workstation that maintains the TSAR system. This is accomplished via the File Transfer Protocol (FTP) either from the UNISYS or any DEC VAX platform. The TSAR development team can aid in this transfer.

The required format for the ASCII input file is as follows:

DATA ITEM	LENGTH	POSITION
DSNAME	8	1 - 8
REFNAME	12	9 - 20
TITLE1	80	21 - 100
TITLE2	80	101 - 180
TITLE3	80	181 - 260
TITLE4	80	261 - 340
DESC	200	341 - 540
SIC	9	541 - 549
GEO	15	550 - 564
UNIT	2	565 - 566
INDEX	4	567 - 570
SEASADJ	1	571 - 571
TRADADJ	1	572 - 572
HOLIDAY	1	573 - 573
BENCHADJ	1	574 - 574
FREQ	4	575 - 578
NOTES	160	579 - 738
PUBLSER	1	739 - 739
PERMSER	1	740 - 740
DERVSER	1	741 - 741

STCKFLOW	1	742 - 742
USRFLAG1	1	743 - 743
USRFLAG2	1	744 - 744
USRFLAG3	1	745 - 745
USRFLAG4	1	746 - 746
USRFLAG5	1	747 - 747
USRFLAG6	1	748 - 748
START	6	749 - 754
END	6	755 - 760
EOR	1	761 - 761

The data items contained in the ASCII input file used to create the Data Descriptor are as follows:

DSNAME ? This data item is the data series name. It contains an eight (8) character name which is the name of the SAS data set which will be used to store the actual time series vector. This data item can be composed of both numerics and characters, it usually is a combination of both types of characters.

This data item cannot contain any special characters, it must contain only numerics (0 -9) and alphabetics (a-z, A-Z). It is not case sensitive. In addition , the first character must be an alphabetic (a-z, A-Z).

The suggested nomenclature for series names is as follows:

CHARACTERS	DESCRIPTION
1	Series Content Indicator
	For example:

U - unadjusted series
A - adjusted series
S - seasonal factor
I - imputation percentage

2 - 6	Classification code, such as SIC
7 - 8	Type of Data Indicator
	For example:

VS for value of shipments, or
TI for Inventories.

REFNAME ? This is the time series reference name. This data item contains a maximum of twelve (12) characters composed of numerics and characters. The reference name is the actual series name used in the prior TimeBase system. It is provided here for historical reference purposes.

TITLE1 - This is a 80 character data item that contains the title of the time series. This title is purely textual and is designed to be used in printouts to describe the time series.

TITLE2 - Additional title information, also a maximum of 80 characters.

TITLE3 - Additional title information, also a maximum of 80 characters.

TITLE4 - Additional title information, also a maximum of 80 characters.

DESC - This is a description of the derivation of the time series. Included in this data item can be the actual derivation equation (algorithm to create this series), or a textual description of the contents of the time series. The intent of this data item is strictly informative and definitional. This data item is not intended to be used in tables or other printed outputs. For series contained in the prior TimeBase system this information contains the derivation equation that was stored for each series by that system. The maximum length of this data item is 200 characters.

SIC - This data item contains a numeric code that classifies the contents of the series. Typically, this is the Standard Industrial Classification code associated with this time series. This may be an SIC code at any digit of aggregation level from a seven?digit product code to a two?digit major group. If the classification scheme for the series is not SIC based then this field can contain any code appropriate for the time series. This field is nine (9) characters in length.

GEO - This data item contains the geographic code (if applicable) for the time series. This field is fifteen (15) characters in length. The standard format for these fifteen characters is as follows:

CHARACTERS	DESCRIPTION
1	Region Code
2	Division Code
3 - 4	State Code (FIPS State Code)
5 - 8	MSA, CMSA, or PMSA
9 - 11	County Code
12 - 15	Place Code

Nonapplicable characters can be zero or space filled.

UNIT - This data item contains the unit?of?measure for the data value for the time series. It defines what is being measured (quantified) by the series. This field contains a maximum of two (2) ASCII characters. The allowable data values are as follows:

1 ? Dollars ($)
2 ? Hundreds of Dollars ($100)
3 ? Thousands of Dollars ($1,000)
4 ? Millions of Dollars ($1,000,000)
5 ? Combined Adjustment Factors (usually in decimal notation)
6 - Seasonal Adjustment Factors (usually in decimal notation)
7 - Trading Day Adjustment Factors (usually in decimal notation)
8 - Holiday Adjustment Factors (usually in decimal notation)

9 - Benchmark Adjustment Factors (usually in decimal notation)
10 - Other Adjustment Factors
11 ? Ratios (usually in decimal notation)
12 ? Quantity (number) of outputs, a measure of physical quantities of economic output (usually in integer notation)
13 ? Quantity in thousands of units
14 ? Quantity in millions of units
15 ? Indexed data
16 - Standard Errors
17 - Coefficients of Variation
18 ? Other

INDEX - This data item contains the base period year for the series. This information is stored as a date field in the data descriptor file. Many economic time series are indexed so that the data represent relative measures to some base time period. In these indexed time series the data value for the base period is usually expressed as one (1). This field is typically blank for most series. This data field contains four (4) ASCII characters in the format YYYY. Where YYYY represents a year, such as 1958 or 1972.

SEASADJ - This data item is used to indicate if the time series is seasonally adjusted. Time series contained within this system can be seasonally adjusted or unadjusted. The allowable one character data values for this data item are as follows:

0 ? Unknown
1 ? Data are NOT SEASONALLY adjusted
2 ? Data ARE seasonally adjusted
3 - Not applicable

TRADADJ - This is the Trading Day Adjustment indicator. The allowable one character data values for this data item are as follows:

0 ? Unknown
1 ? Data are NOT adjusted for trading day factors

2 ? Data ARE adjusted for trading day factors
3 ? Not Applicable

HOLIDAY - This is the Holiday Adjustment indicator. The allowable one character data values for this data item are as follows:

0 ? Unknown
1 ? Data are NOT adjusted for holiday factors
2 ? Data ARE adjusted for holiday factors
3 ? Not Applicable

BENCHADJ - This is the Benchmark Adjustment indicator. The allowable one character data values for this data item are as follows:

0 ? Unknown
1 ? Data are NOT adjusted by benchmark factors
2 ? Data ARE adjusted by benchmark factors
3 ? Not Applicable

FREQ - This data item defines the frequency of the series. This data value indicates the number of data points included within the series for a year.

Typical frequencies entries are:

ENTRY	DESCRIPTION
1	Annual data
2	Semi-Annual
4	Quarterly data
6	Bi-Monthly
12	Monthly data
52	Weekly data
365	Daily Data

NOTES - This item is provided to store any descriptive textual information required by the system user. The field width for NOTES is 160 characters.

PUBLSER - Published series indicator. The allowable data values are as follows:

0 - Unknown
1 - Time series is NOT published
2 - Time series IS published

PERMSER - Permanent series indicator. The allowable data values for this item are as follows:

0 - Unknown

1 - Temporary series
2 - Permanent series

DERVSER - Derived series indicator, this indicates that the series is mathematically derived (calculated). The data values in this series are algorithmically determined. The allowable data values are as follows:

0 - Unknown
1 - Series is NOT derived, data are directly loaded into the series
2 - Derived series

STCKFLOW - Stock or flow indicator. The allowable data values are:

0 - Unknown
1 - Stock Series
2 - Flow Series

USRFLAG1 - This item is provided for user defined flags.

USRFLAG2 - This item is provided for user defined flags.

USRFLAG3 - This item is provided for user defined flags.

USRFLAG4 - This item is provided for user defined flags.

USRFLAG5 - This item is provided for user defined flags.

USRFLAG6 - This item is provided for user defined flags.

START - This is the beginning date of the time series. This field is six (6) characters in length. The format is YYYYFP ; where YYYYFP is interpreted as follows:

YYYY - represents the year, such as 1958 or 1967;

FP - represents the frequency (F) and the period (P);

The allowable values for various frequencies are:

A - annual data
Q - quarterly data
N - biennial data
S - supplemental data
B - biannual data
T - triannual data
U - quinquennial data

For monthly data the frequency is not used, the period is provided as two characters. For all other (non-monthly) frequencies the period is provided as a single character.

If monthly data the allowable period values are: 01 - 12, for example 199509 is September 1995;

If quarterly data the allowable frequency and period values are: Q1 -Q 4, for example 1994Q2 is the second quarter of 1994;

END - This is the ending date of the time series. This field is six (6) characters in length. The format is YYYYFP ; where YYYYFP is interpreted as follows:

YYYY - represents the year, such as 1958 or 1967;

FP - represents the frequency (F) and the period (P);

The allowable values for various frequencies are:

A - annual data

Q - quarterly data
N - biennial data
S - supplemental data
B - biannual data
T - triannual data
U - quinquennial data

For monthly data the frequency is not used, the period is provided as two characters. For all other (non-monthly) frequencies the period is provided as a single character.

If monthly data the allowable period values are: 01 - 12, for example 199509 is September 1995;

If quarterly data the allowable frequency and period values are: Q1 -Q 4, for example 1994Q2 is the second quarter of 1994;

The output of this function is a SAS data set named: descript.ssd01.